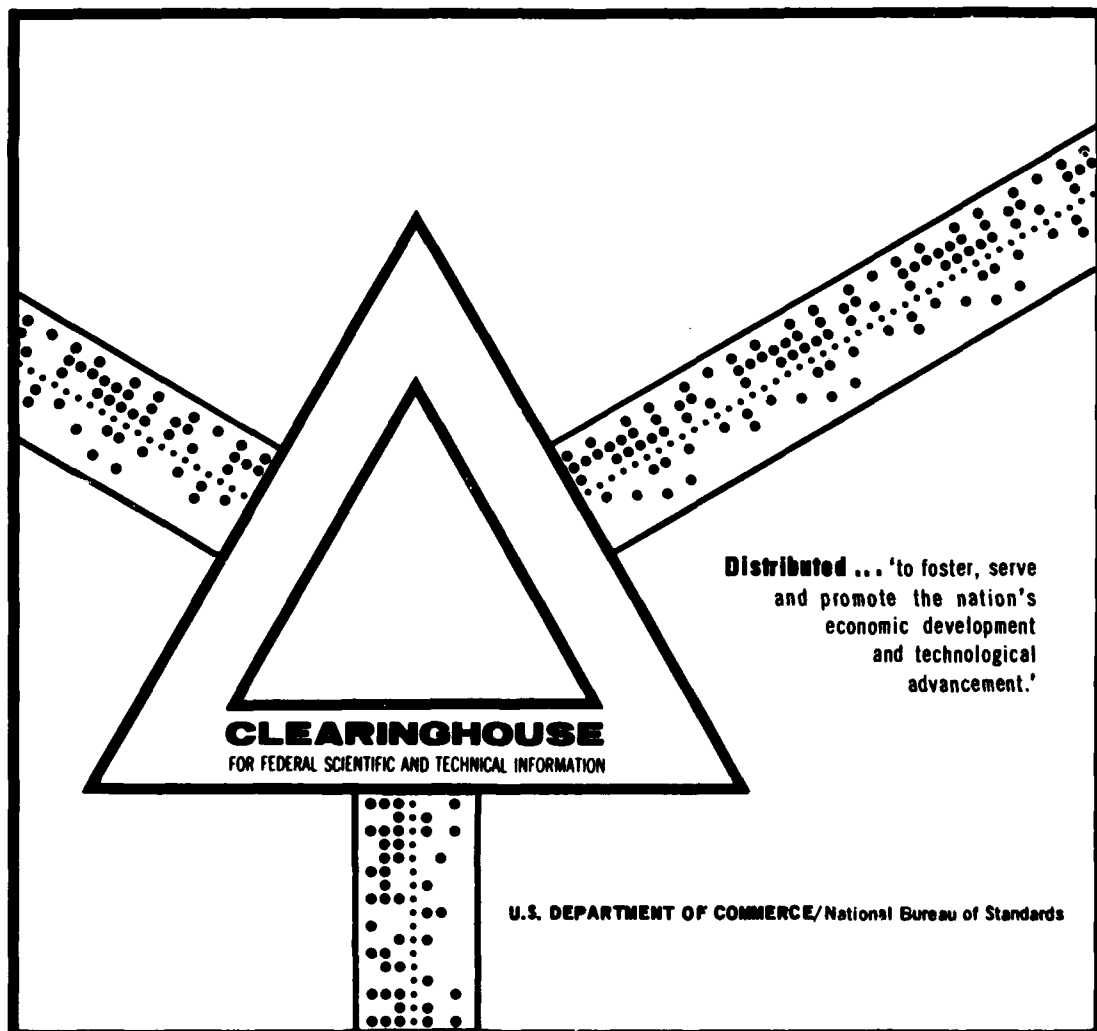# AN ALGORITHM FOR NONCONVEX PROGRAMMING

G. Graves, et al

Purdue University
Lafayette, Indiana

September 1969

Distributed ... 'to foster, serve
and promote the nation's
economic development
and technological
advancement.'

CLEARINGHOUSE
FOR FEDERAL SCIENTIFIC AND TECHNICAL INFORMATION

U.S. DEPARTMENT OF COMMERCE/National Bureau of Standards

AN ALGORITHM FOR NONCONVEX PROGRAMMING[*]

by

G. GRAVES
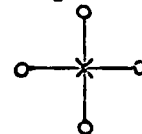
and

A. WHINSTON

AN ALGORITHM FOR NONCONVEX PROGRAMMING

G. Graves and A. Whinston

1. Introduction

This paper presents an algorithm to solve the most general mathematical programming problem

$$\text{s.t. } g^i(y) \le 0 \qquad i = 1,2,\ldots,m$$

$$\text{Min. } g(y) \qquad y = (y_1,\ldots,y_n)$$

The only restriction required is that the functions g^i, g be real valued. The general formulation allows for nonlinear or linear integer programming, mixed integer programming and general nonconvex continuous variable programming. The extant algorithms for this most general problem can usually be viewed as local search procedures. They suffer from two serious difficulties which can be characterized as the "dimensionality problem" and the problem of "trapping at local optima". These difficulties are illustrated by the "local corner search" where each of the 2^n adjacent corners of a current point are evaluated and the best of these is used as the next current point. The number of function evaluations increase exponentially with the number of variables and the procedure is impossible except for problems with very few variables. As is well known, this procedure stabilizes at local optima. Traditionally, convexity is invoked by mathematicians to eliminate this sort of unpleasantness. As a practical matter
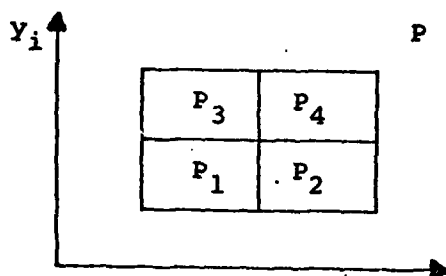
with real problems, convexity is never established. In fact, the essence of location of facilities problems is precisely the tradeoff between the economies of scale in production and the transportation cost. (Economies of scale imply minimization of concave functions.)

The classical approaches, then, have been essentially "local" or "neighborhood" techniques dependent on derivatives (or finite difference approximations to derivatives). Only unrealistic assumptions such as "convexity" or vague arm waving such as "try a representative sample of starting points" have been advocated to deal with the global problem. (Obtaining a "representative sample of starting points" is feasible with small generally artificial examples.) We feel this sweeps the very quintessence of many economic problems under the rug. Our central aim here is to present a new framework for reaching global optimum. The procedure involves two interconnected mechanisms, a method for structuing the search and a decision rule for selecting the course of the search.

2. Structuring the Search

Structuring the search consists of introducing a framework for reducing the general problem to that of "implicit enumeration" [1]. In general, given a bounded domain P, it can be symetrically partitioned into subdomains $P_1, P_2, \ldots, P_2n$. For example,

Technically:

$$\text{given} \quad b(i) \leq y(i) \leq S(i)$$

$$\text{define} \quad r(i) = (S(i) - b(i))/2$$

$$\overline{y}(i) = b(i) + r(i)$$

and introduce the class C of finite maps

$$\omega: \{1,\ldots,n\} \to \{0,1\}.$$

Now a 1-1 correspondence can be setup between the components P_i of the partition of P and the class of maps C by defining the upper and lower bounds of a component in terms of a map ω,

$$L(i,\omega(i)) = \overline{y}(i) - (1-\omega(i)) \cdot r(i)$$

$$U(i,\omega(i)) = \overline{y}(i) + \omega(i) \cdot r(i)$$

To illustrate these formulas, we can apply them to the two dimensional unit square. In this event,

$$0 \leq y(i) \leq 1 \qquad i = 1,2$$

$$\text{e.g.} \quad b(1) = 0 \qquad\qquad S(1) = 1$$

$$b(2) = 0 \qquad\qquad S(2) = 1$$

$$\text{and} \quad r(1) = 1/2 \qquad\qquad r(2) = 1/2$$

$$\overline{y}(1) = 1/2 \qquad\qquad \overline{y}(2) = 1/2$$

Using these quantities:

$$L(1,\omega(1)) = 1/2 - (1-\omega(1))\cdot 1/2$$
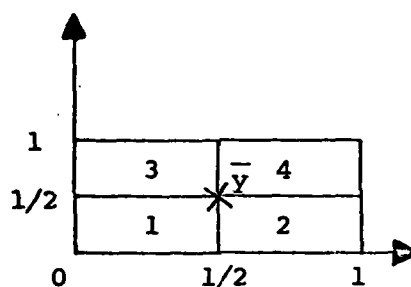
$$U(1,\omega(1)) = 1/2 + \omega(1)\cdot 1/2$$

and

$$L(2,\omega(2)) = 1/2 - (1-\omega(2))\cdot 1/2$$

$$U(2,\omega(2)) = 1/2 + \omega(2)\cdot 1/2.$$

The choice of any of the four different maps $(\omega(1),\ \omega(2))$ specifies a particular rectangle.

For example, consider the map $(0,1)$, e.g.,

$$\omega(1) = 0$$

$$\omega(2) = 1$$

This map specifies rectangle 3,

$$L(1,\omega(1)) = 0 \qquad U(1,\omega(1)) = 1/2$$

$$L(2,\omega(2)) = 1/2 \qquad U(2,\omega(2)) = 1.$$

The problem is now reduced to choosing a desirable map ω_* and further refining the corresponding component until a point is specified to any predetermined accuracy.

Technically this can be setup recursively by taking,

$$r^{o}(i) = (S(i) - b(i))/2$$

$$\overline{y}^{o}(i) = b(i) + r^{o}(i)$$

and

$$r^t(i) = r^{(t-1)}(i)/2$$

$$\bar{y}^t(i) = \bar{y}^{(t-1)} - (1-\omega_*^{(t-1)}(i)) \cdot r^t(i) + \omega_*^{(t-1)}(i) \cdot r^t(i)$$

and again for any map ω at the ith stage,

$$L^t(i,\omega(i)) = \bar{y}^t(i) - (1-\omega(i)) \cdot r^t(i)$$

$$U^t(i,\omega(i)) = \bar{y}^t(i) + \omega(i) \cdot r^t(i).$$

Now specifying a sequence of maps

$$(\omega_*^0, \omega_*^1, \omega_*^2, \ldots)$$

specifies a sequence of nested intervals for each i

$$[L^t(i,\omega_*^0(i)), U^t(i,\omega_*^0(i))]$$

such that

$$L^t(i,\omega_*^t(i)) \longrightarrow \text{(monotonically increases with t)}$$

$$U^t(i,\omega_*^t(i)) \longrightarrow \text{(monotonically decreases with t)}$$

and $\quad [U^t(i,\omega_*^t(i)) - L^t(i,\omega_*^t(i))] = (S(i) - b(i))/2^{t+1}$

$\rightarrow 0$ as $t \rightarrow \infty$.

Therefore a sequence of maps $(\omega_*^0, \omega_*^1, \ldots)$ defines

an n-tuple of real numbers or a point in R^n. (Recall the

Weirstrauss_Heine development of the real numbers. Their

definition is: "A real number is a nest of intervals (x_n, y_n)

such that $\{x_n\}$ is monotonic increasing, $\{y_n\}$ is monotonic

decreasing, and $d_n = (y_n - x_n) \rightarrow 0$ as $n \rightarrow \infty$." See Knopp

[3], Chapter 1.) Now with any stipulated accuracy for the solution

$y^*(i) \triangleq \epsilon$ take the first positive integer T such that $(S(i) - b(i))/2^{T+1}$

$\leq \epsilon$ for all i or $2^{T+1} \geq (S(i) - b(i))/\epsilon$.

Introduce the class $\bar{C}$ of meta-maps

$$\phi: \{1,2,\ldots,nxT\} \to \{0,1\}.$$

The choice of a ϕ determines a "quantitized" point in the
domain of interest. The problem is reduced to choosing the
optimal meta-map ϕ^*. The algorithm we propose is to <u>implicitly
enumerate</u> the class $\bar{C}$ of meta-maps. There are of course many
other ways of "quantitizing" the domain suitable for implicit
enumeration. The employment of the present structure and,
in particular, the T sub-maps $(\omega^1,\omega^2,\ldots,\omega^T)$ to specify ϕ is
to isolate for easy exploitation the nested components of
the successive partitions identified by the ω^t. <u>It is these
nested components that allow us to introduce set functionals
for decision making and a global approach to calculating
the optimum independent of such restrictions as convexity
on the original functions.</u>

3. Decision Rules for Directing the Search.

The most common set functional in mathematics is the
ordinary integral. It is our contention that use of this
functional instead of resorting to the derivative or its
finite difference counterpart of the "local" procedures
should enable us to utilize global information. Liberating
our decision process from the myopic local neighborhood
processes should render us insensitive to trapping at local
optima and enable us to dispense with inapplicable mathemati-
cal assumptions such as "convexity". The most elementary

use of the integral would be to simply calculate for each component (defined by an element ω^t of the meta-map) the following quantities:

$$AV(\omega^t,g) = \frac{1}{\prod_i r(i)} \int_{L^t(1,\omega(1))}^{U^t(1,\omega(1))} \cdots \int_{L^t(n,\omega(n))}^{U^t(n,\omega(n))} g(y)\,dy_1,\ldots dy_n$$

$$SS(\omega^t,g) = \frac{1}{\prod_i r(i)} \int_{L^t(1,\omega(1))}^{U^t(1,\pi(1))} \cdots \int_{L^t(n,\omega(n))}^{U^t(n,\omega(n))} g^2(y)\,dy_1\ldots dy_n$$

$$SGM(\omega,g) = [(SS(\omega^t,g) - AV^2(\omega^t,g)]^{1/2}$$

$$d(\omega^t,g) = AV(\omega,g) - v \cdot SGM(\omega,g)$$

The component of the meta-map ω_*^t chosen would be such that $d(\omega_*^t,g) = \min_{\omega \in C} d(\omega^t,g)$.

The decision functional $d(\omega^t,g)$ is a simple estimator of the minimum value of the function $g(y)$ on the associated component of the partition. If no knowledge of the underlying distribution is available, the parameter v in the definition of $d(\omega^t,g)$ would have to be determined empirically or several runs made using various values.

This simple procedure suffers from the same "dimensionality problem" as the local search procedures. The evaluation of the decision functional $d(\omega^t,g)$ for all possible 2^n maps ω^t would impose an intolerable computational burden (except for artificial mathematical examples). This "dimensionality problem" can be eliminated, however, by

resorting to an n-stage sequential decision process. The total map ω^t would be constructed in n-steps by sequentially fixing elements of the map. Suppose an arbitrary set of k out of the possible n elements of the domain are fixed. At the $(k+1)^{st}$ step an additional element of the domain, say ℓ_{k+1} is choosen and

$$\ell_{k+1} \to 0 \quad \text{or} \quad \ell_{k+1} \to 1.$$

Now if the order of fixing elements of the domain is completely arbitrary, there would be $2(n-k)$ possible choices of a couple $(\ell_{k+1}, 0)$ or $(\ell_{k+1}, 1)$ at each stage. The total number of functional evaluations would reduce to

$$\sum_{k=0}^{n-1} 2(n-k) = 2 \sum_{k=1}^{n} k = n(n+1).$$

(This reduction is insignificant for 3 or 4 variables, but with as few as 20 variables we would achieve a reduction from

$$2^{20} = 1,048,576$$

to

$$20 \cdot 21 = 420.)$$

In the n-stage sequential process, it is necessary to use a slightly more sophisticated decision functional. Each choice is now determined by expected values over all completions of the k-partial map. Given a k-partial map, for example,

$$\begin{pmatrix} \overset{\text{fixed}}{1} & 2 & \ldots & k \enspace \Big\} & \overset{\text{free}}{k+1} & \ldots & n \\ 0 & 1 & \ldots & 1 \enspace \Big\} & 0 & \ldots & 1 \end{pmatrix}$$

there are 2^{n-k} possible completions. We then employ the following expected values over the completion class C_k

$$E_{C_k}(AV(\omega^t,g)) = \frac{1}{2^{n-k}\pi_r(j)} \int_{L^t(1,\omega(1))}^{U^t(1,\omega(1))} \cdots \int_{L^t(k-1,\omega(k))}^{U^t(k,\omega(k))}$$

$$\cdots \int_{L^t(k+1,0)}^{U^t(k+1,1)} \cdots \int_{L^t(n,0)}^{U^t(n,1)} g(y)dy_1 \ldots dy_n$$

$$E_{C_k}(SS(\omega^t,g)) = \frac{1}{2^{n-k}\pi_r(j)} \int_{L^t(1,\omega(1))}^{U^t(1,\omega(1))} \cdots \int_{L^t(k,\omega(k)}^{U^t(k,\omega(k))}$$

$$\cdots \int_{L^t(k+1,0)}^{U^t(k+1,1)} \cdots \int_{L^t(n,0)}^{U^t(n,1)} g^2(y)dy_1 \ldots dy_n.$$

These results, of course, rely on the "additivity" of the limits of integration.

Using these more sophisticated quantities we proceed as before by calculating

$$SGM^{(k)}(_\omega{}^t,g) = [E_{C_k}(SS(\omega^t,g)) - E^2_{C_k}(AV(\omega^t,g))]^{1/2}$$

that is, the standard deviation of $g(y)$ on the components and

$$d^{(k)}(_\omega{}^t,g) = E_{C_k}(AV(\omega^t,g)) - v \cdot SGM^{(k)}(\omega,g).$$

The decision functional $d^{(k)}(_\omega{}^t,g)$ is evaluated for the $2(n-k)$ possible couples, say $(\ell_{k+1}) \rightarrow 1$ or $(\ell_{k+1},0)$ and ℓ_{k+1} any

"free" element of the k-partial map. The minimum value of $d^{(k)}(\omega^t, g)$ determines the next couple to be fixed.

This whole n-stage sequential decision process is then carried out T times as indicated in Section 1 to yield a "point" in R^n which is hopefully very close to the global minimum of $g(y)$. In any event, by continuing and employing a "confidence level implicit enumeration" (see [1] and [2]) of the whole class $\overline{C}$ of meta-maps, we should achieve a highly sophisticated search of the whole domain. The only point to note in employing the mechanism of the "confidence level enumeration" is that the recursive definition of the components would require ω^i to be entirely fixed before any element of ω^{i+1}.

4. Additional Observations

(A) Limiting Value

When the function $g(y)$ is continuous, it might be worth noting that

$$d^{(n)}(\omega^t, g) \rightarrow g(y^*) \quad \text{as } t \rightarrow \infty$$

where $y^* \in R^n$

is the point defined by the sequence of maps $(\omega_*^0, \omega_*^1, \ldots)$. This follows immediately from the Mean Value Theorem for Integrals which says:

$$\frac{1}{AREA} \int_D g\,dA = g(\tilde{y})$$

where $\tilde{y} \in D$.

Applying this result to the terms of $d^{(n)}(\omega^t, g)$ yields

$$E_{c_n}(AV(\omega^t, g)) \to g(y^*) \qquad \text{as } t \to \infty$$

$$SGM^{(n)}(\omega^t, g) \to 0 \qquad \text{as } t \to \infty$$

and hence

$$d^{(n)}(\omega^t, g) \to g(y^*) \qquad \text{for any } v \text{ as } t \to \infty.$$
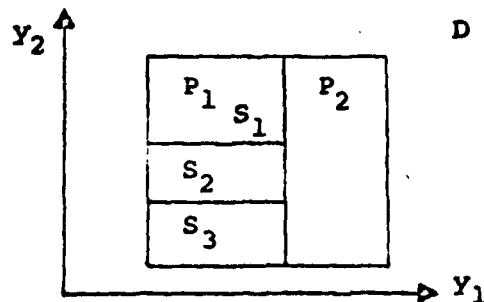
(B) Indefinite Integral

The evaluation of the integrals employed in the definition of the decision functional $d^{(k)}(\omega^t, g)$ can be carried out in various ways. With continuous functions, the simplest procedure is to employ the closed form given by the indefinite integral; for example,

$$\frac{1}{AREA} \int_D y_1 y_2 \, dy_1 dy_2 = \frac{(U_1^2 - L_1^2) \cdot (U_2^2 - L_2^2)}{4(U_1 - L_1) \cdot (U_2 - L_2)} = \frac{(U_1 + L_1) \cdot (U_2 + L_2)}{4}$$

(C) Stratified Sampling

When the function is not known in closed form or the indefinite integral is not available, it may become necessary to resort to stratified sampling of the various components of the domain defined by the limits of integration in the decision functional; for example,

(Insure uniform distribution of points over component.)

We could determine an appropriate sample size k_i for each strata S_i and on the basis of this sample calculate estimates

$$\hat{E}_{C_k}(AV(\omega^t, g)) \qquad \text{and} \qquad \hat{SGM}^{(k)}(\omega^t, g)$$

and from these calculate $d^{(k)}(\omega^t, g)$. At any decision point, we are stratifying a domain of the form:

$$
\begin{aligned}
L^t(1, \omega(1)) &\le y_1 &\le U^t(1, \omega(1)) \\
&\vdots& \\
L^t(k, \omega(k)) &\le y_k &\le U^t(k, \omega(k)) \\
L^t(k+1, 0) &\le y_{k+1} &\le U^t(k+1, 1) \\
&\vdots& \\
L(n, 0) &\le y_n &\le U^t(n, 1) \ .
\end{aligned}
$$

(D) Discrete Variables

It is, of course, not necessary that the variables be continuous. The Rieman-Stieltzes Integral is available to deal with discrete variables. Recall the usual Unit Step Function

$$
I(y) = \begin{cases} 0 & (y \le 0) \\ 1 & (y > 0) \end{cases}
$$

and Standard Counting Measure

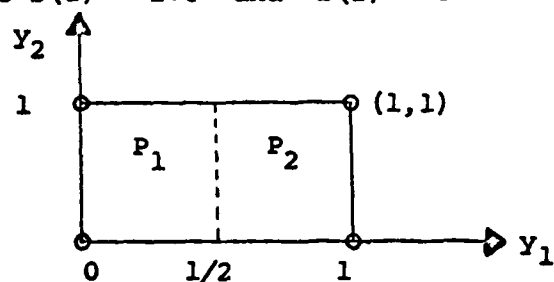$$\alpha_i(y_i) = I(y_i) + I(y_i - 1)$$

that would be employed with zero-one discrete variables. (A slight generalization would eliminate the reduction to zero-one discrete variables.) In this formulation,

$$\text{"AREA"} = \int_0^{1+\epsilon} \ldots \int_0^{1+\epsilon} d\alpha_1 \ldots d\alpha_n = 2^n$$

and for illustrative purposes, consider the simple linear case

$$g(y) = \sum_{i=1}^{n} a_i y_i.$$

Take $S(i) = 1+\epsilon$ and $b(i) = 0$

"Mass Points"

(0,0), (0,1), (1,0), (1,1)

$\underline{P_1}$ $\quad E_{C_k}(AV(\omega^t,g)) = \frac{1}{2} \int_0^{1+\epsilon} \int_0^{\frac{1}{2}} (a_1 y_1 + a_2 y_2) d\alpha_1 d\alpha_2$

$$= \frac{1}{2} \int_0^{1+\epsilon} a_2 y_2 d\alpha_2 = \frac{a_2}{2}$$

$\underline{P_2}$ $\quad E_{C_k}(AV(\omega^t,g)) = \frac{1}{2} \int_0^{1+\epsilon} \int_{1/2}^{1+\epsilon} (a_1 y_1 + a_2 y_2) d\alpha_1 d\alpha_2$

$$= \frac{1}{2} \int_0^{1+\epsilon} (a_1 + a_2 y_2) d\alpha_2 = a_1 + \frac{a_2}{2}$$

Hence, as expected, the decision of whether $y_1 \to 0$ or $y_1 \to 1$ is determined by whether $a_1 < 0$ or $a_1 > 0$. This general approach reduces to techniques expounded in great detail in [1]. It should be stressed that the Rieman-Stieltzes Integral Approach developed in this section is perfectly capable of handling pure continuous variables, mixed continuous and integer variables, or pure integer variables.

(E) Constraints

The ideas developed in this paper can be extended to treat constraints of the form

$$g^i(y) \leq 0 \qquad i = 1,2,\ldots,m$$

by introducing conditional expected values. The simplest way to achieve this is through the use of a Regression Equation. Instead of using $AV(\omega^t,g)$, this would require employment of:

$$AV(\omega^t,(g \mid g^i)) = AV(\omega^t,g) + \frac{COV(\omega^t,g,g^i)}{Var\ (g^i)} \cdot (g^i - AV(\omega^t,g^i)),$$

the conditional expected value of the function g given a value of function g^i. In this procedure, it would be necessary to estimate the maximum or minimum of $(g^i - AV(\omega^t,g^i))$ depending on the sign of the covariance on the components specified by the current k-partial map ω^t. This could be done in turn in terms of the variance of g^i and its mean. It would also be necessary to establish an appropriate confidence level that $g^i(y) \leq 0$ on the component. When the confidence drops too low it is necessary to "backtrack" in the construction of the meta-map. It should be observed that "normality" assumptions are not required for this procedure, but in the event of non-normality, the linear regression equation reduces to a first order approximation. Again, these ideas are developed at greater length in [1].

5. Implementation of the Algorithm

The algorithm has been programmed in FORTRAN IV and several example problems have been run on the C.D.C. 6500. Numerous approaches to implementing the ideas described above are possible. The present program determines indefinite integral and successively reevaluates at the upper and lower bounds for the appropriate domains as specified in the search procedure. This approach limits the size of the problems that can be handled since the storage requirements for the mean and variance are considerable. For large scale problem use of the ideas discussed under stratified sampling or numerical approximations may prove more effective. It should be noted that the decision process is generally insensitive to the errors resulting from an approximate determination of the values of the decision variables.

Several problems obtained from the literature were used to test the algorithm. Each problem was run for a succession of values of v_j for $v = .5$, $.15$, 1 and 1.5. Below we list some results obtained using a C.D.C. 6500. Since the code was experimental a considerable amount of time was taken in calculation not directly used, including the determination of higher moments. A rough estimate would be that the problems could be solved in one half of the stated time. Further precision could be obtained in the answer by using a gradient method in the vicinity of the solution obtained by this algorithm.

Problem 1   Min $\frac{1}{2}x_1 - \frac{1}{2}x_2 - \frac{1}{2}x_1^2 + \frac{1}{2}x_2^2$

s.t. $2x_1 + x_2 \leq 6$

$-x_1 + 4x_2 \leq 6$

$x_i \geq 0 \quad i = 1,2$

The optimal solution is (3, 0) and a value of -3 for the criterion function. We obtained (2.929, .07812) and -2.8627 with $v = 1.5$ in 5 seconds of computer time.

Problem 2   Max $x_1 \, x_2 \, x_3$

$x_1 + 2x_2 + 2x_3 \leq 72$

$0 \leq x_i \leq 42 \quad i = 1,2,3$

The answer is (24, 12, 12) with a value of 3, 456 and we obtained (25.02) 9.844, 13.617) with a value of 3, 354. The value of v was 15 and the computations took about 5 seconds.

REFERENCES

[1] Graves, G. and A. Whinston, "A New Approach to Discrete Mathematical Programming," *Management Science*, Nov. 1968.

[2] Graves, G. and A. Whinston, "An Algorithm for the Quadratic Assignment Problem," forthcoming.

[3] Knopp, K., *Elements of Functions*, Dover Publications.

DOCUMENT CONTROL DATA - R&D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author)	2a. REPORT SECURITY CLASSIFICATION
Krannert Graduate School of Industrial Administration Purdue University Lafayette, Indiana 47907	2b. GROUP

3. REPORT TITLE

An Algorithm for Nonconvex Programming

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

Technical Report

5. AUTHOR(S) (Last name, first name, initial)

G. Graves and A. Whinston

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
September 5, 1969	16	3

8a. CONTRACT OR GRANT NO. N00014-67-A-0226-0006 MOD #AC	9a. ORIGINATOR'S REPORT NUMBER(S)
b. PROJECT NO. 5157-53-1310	
c.	9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)
d.	

10. AVAILABILITY/LIMITATION NOTICES

Distribution of this document is unlimited

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY Logistics and Mathematical Science Branch Office of Naval Research, Washington, D.C.

13. ABSTRACT This paper presents an algorithm to solve the most general mathematical programming problem:

$$\text{s.t.} \quad g^i(y) \leq 0 \qquad i = 1,2,\ldots,m$$
$$\text{Min.} \quad g(y) \qquad y = (y_1,\ldots,y_n).$$

The only restriction required is that the functions g^i, g be real valued. The general formulation allows for nonlinear or linear integer programming, mixed integer programming and general nonconvex continuous variable programming.

The classical approaches have been essentially "local" or "neighborhood" techniques dependent on derivatives (or finite difference approximations to derivatives). They suffer from two serious difficulties which can be characterized as the "dimensionality problem" and the problem of "trapping at local optima."

Our central aim here is to present a new framework for reaching global optimum. The procedure involves two interconnected mechanisms, a method for structuring the search and a decision rule for selecting the course of the search.

DD FORM 1473